

ICOLC 2006: SUSHI un nuovo strumento per la misurazione delle risorse elettroniche

Domenico Dellisanti

CILEA, Segrate

Abstract

Relazione sulla sessione delle statistiche d'uso all'interno del convegno ICOLC 2006 (Roma, ottobre).

Report about usage statistics session of the ICOLC 2006 meeting. (Rome, october).

Keywords: ICOLC, usage statistics, SUSHI, COUNTER, web services, ERM.

Si è tenuto a Roma, dal 11 al 14 ottobre 2006, l'ottavo convegno annuale dell'ICOLC (International COalition of Library Consortia) [1] sponsorizzato dal CILEA [2], dal CASPUR [3] e dall'Università degli Studi di Roma 3 [4].

Durante il convegno sono state affrontate tematiche di notevole interesse per i consorzi che si occupano di risorse elettroniche.

Un'intera sessione, denominata "Measuring use and services", è stata dedicata alle statistiche d'uso, tematica che sta assumendo un ruolo sempre più importante nel mondo delle biblioteche digitali, in quanto fondamentale per un buon uso delle risorse e per l'ottimizzazione degli investimenti.

In questa sessione sono stati presentati nuovi prodotti e illustrati progetti e attività di vari consorzi, come quelli dell'Ontario (OCUL) [5], dell'Anatolia (ANKOS) [6] e della Finlandia (FinElib) [7].

La presentazione delle attività svolte da realtà consortili molto diverse tra di loro, sia per distribuzione geografica che per disponibilità finanziaria, permette di avere una visione di come gli altri consorzi nel mondo lavorano e cercano di ottenere buoni risultati, in termini di utilizzo delle proprie risorse elettroniche.

All'interno della sessione dedicata alle statistiche d'uso, una relazione che ha catturato l'attenzione dei presenti è stata quella che ha illustrato l'iniziativa SUSHI [8], acronimo di Standardized Usage Statistics Harvesting Ini-

ziative presentata, da Oliver Pesh della EBSCO Information Services.

SUSHI non si contrappone a COUNTER (Counting On-line User NeTworked Electronic Resources) [9], standard utilizzato dalla maggior parte degli editori in questi ultimi anni per la produzione delle statistiche d'uso, ma ne utilizza i dati, al fine di mettere a disposizione informazioni facilmente interpretabili e quindi di grande utilità.

Perché SUSHI: breve storia

Come accennato prima, negli ultimi anni, quando si parlava di elaborazione di statistiche d'uso nel campo delle risorse elettroniche, si faceva riferimento principalmente allo standard COUNTER.

COUNTER fornisce un ottimo modello per la produzione di statistiche d'uso, ma non soddisfa totalmente i bisogni della comunità.

Uno dei problemi che i bibliotecari si trovano ad affrontare è che i rapporti generati secondo lo standard COUNTER sono disponibili in fogli Excel, con i dati provenienti da ciascun editore elaborati su file differenti ossia poco aggregabili. La soluzione più naturale è stata quella di pensare a un sistema che accettasse una richiesta ed emettesse una risposta seguendo un protocollo basato sui servizi web e sugli schemi XML.

Le origini di SUSHI risalgono al novembre del 2004, quando ci si incominciò a porre interrogativi sull'utilizzo delle statistiche d'uso elaborate secondo lo standard COUNTER da parte

dei moduli presenti nei vari ERM (Electronic Resource Management). Un ERM può essere definito come un software che permette a una biblioteca di gestire le risorse elettroniche a sua disposizione, oltre che le sottoscrizioni cartacee siglate con i vari editori. Questa gestione può coprire tutte le fasi del ciclo di vita di una risorsa (accesso, acquisizione, eventuale periodo di prova, analisi dei costi, pagamenti, dati statistici). I bibliotecari necessitavano di uno modello di scambio di dati più efficiente.

In particolare non esistevano dei metodi di aggregazione degli stessi dati e il processo di questa operazione era spesso molto lungo e faticoso. Si è quindi sentito il bisogno di studiare e realizzare un modello più efficiente, che venisse incontro alle esigenze di chi doveva recuperare e, successivamente, analizzare questi dati.

Nel 2005, si diede vita al primo comitato, in cui entrarono a far parte membri appartenenti a enti accademici (Cornell University, University of Washington, Harvard University), aggregatori (EBSCO), editori (Thomson Scientific) e case produttrici di ERM (Ex Libris, Innovative Interfaces). Successivamente, SUSHI divenne un'iniziativa NISO (National Information Standards Organization) [10].

Attualmente, ai membri fondatori se ne sono aggiunti altri, tra cui ricordiamo Endeavour Information Systems, Florida Center for Library Automation, College Center for Library Automation, Otto Harrassowitz, OCLC, Project Euclid, Serial Solutions e SirsiDynix.

Nell'estate del 2006 fu sviluppato e testato il primo prototipo, denominato SUSHI 0.1, con la collaborazione della EBSCO, della Swets e di Project Euclid.

Recentemente, SUSHI 0.1 si è evoluto in SUSHI 1.0 e la stessa NISO, con COUNTER, ha siglato un accordo (Memorandum of Understanding) in base al quale SUSHI adopererà gli schemi XML già utilizzati in COUNTER.

Cos'è SUSHI

SUSHI può essere definito un modello basato su un servizio web per la richiesta di dati statistici. L'utente non è più obbligato a fare il download dei dati relativi alle statistiche d'uso dai siti dei vari editori, perché questo compito, ora, viene svolto dallo stesso SUSHI.

I dati che SUSHI acquisisce sono sempre nel formato COUNTER, di cui sono stati conservati gli schemi XML.

Bisogna comunque ribadire che SUSHI non è un modello che permette calcoli su statistiche d'uso.

La tecnologia

Si è scelta la tecnologia basata sui servizi web, in quanto questi combinano molto bene l'ambiente WEB e l'approccio della programmazione modulare. Sono diffusi a livello commerciale, sono supportati dal W3C e permettono di avere un alto livello di sicurezza sul trasferimento dei dati, per garantire la privacy.

Secondo la definizione data dal W3C, un servizio web è un sistema software progettato per supportare l'interoperabilità tra diversi elaboratori su di una medesima rete; caratteristica fondamentale di un servizio web è quella di offrire un'interfaccia software (descritta in un formato automaticamente elaborabile quale, per esempio, il WSDL) utilizzando la quale altri sistemi possono interagire con il servizio stesso, attuando le operazioni descritte nell'interfaccia tramite appositi "messaggi" inclusi in una "busta" SOAP (Simple Object Access Protocol): tali messaggi sono, solitamente, trasportati tramite il protocollo http e formattati secondo lo standard XML [11].

Messaggi di richiesta e di risposta

Il protocollo SUSHI è molto semplice, in quanto supporta solo due tipi di messaggi: messaggio di richiesta e messaggio di risposta. Nel messaggio di richiesta sono presenti tre informazioni fondamentali:

- requestor,
- customerReference,
- reportDefinition.

Il requestor nel file XML è caratterizzato da un campo identificativo univoco, dal proprio nome e dalla propria email.

Il customerReference sempre da un campo id e dal suo nominativo.

Nel reportDefinition sono presenti le informazioni sui dati che si richiedono ossia la tipologia di rapporto e la data di riferimento con possibile data di inizio e data di fine del prescelto periodo per il quale si stanno richiedendo i dati.

Nel messaggio di risposta sono presenti le seguenti informazioni:

- requestor,
- customerReference,
- reportDefinition,
- reports.

Il requestor nel file XML è caratterizzato sempre dai campi: id, name e email.

Nel customerReference ritroviamo le informazioni su chi ha inoltrato la richiesta.

Nel reportDefinition la tipologia di rapporto (per esempio J1), l'intervallo temporale per cui

si stanno richiedendo i dati, con specificata data di inizio e di fine del periodo.

Nel *tag report* sono presenti tutte le informazioni relative al rapporto che è stato elaborato.

Troviamo le informazioni sul *vendor* (nominativo, sito web, email, logo), sul *customer* (nominativo, identificativo) e i dati statistici che sono stati elaborati. Per esempio, nome del giornale, l'online ISSN, il *print* ISSN, il *publisher*, la piattaforma su cui è accessibile il giornale, la data di inizio del periodo per il quale abbiamo richiesto le statistiche, la data di fine, la tipologia di dati restituita (pdf, html, total).

Transazione SUSHI

Vediamo quali sono i passi necessari di una transazione SUSHI:

- una biblioteca richiede un *usage report*,
- il SUSHI Client effettua la richiesta,
- il SUSHI Server processa la richiesta,
- il SUSHI Server prepara i rapporti COUNTER richiesti,
- il SUSHI Server impacchetta i rapporti e prepara la risposta,
- il SUSHI Client riceve la risposta e la processa.

Gli attori in questo processo sono principalmente, da un lato la biblioteca, da cui parte la richiesta, e dall'altro il *provider*, che deve reperire i dati e fornire la risposta.

La biblioteca effettua la richiesta tramite il SUSHI Client, che è un software che gira sul server della biblioteca stessa e solitamente è associato a un ERM. Dal lato provider abbiamo il SUSHI Server, che riceve la richiesta e che reperisce i dati statistici. Questi dati statistici vengono elaborati secondo lo standard COUNTER, impacchettati tramite il SUSHI Server che, una volta terminata l'elaborazione, è in grado di rispondere alla richiesta che gli era stata posta dal SUSHI Client.

Il SUSHI Client riceve la risposta, che consiste in dati statistici che sono *compliant* COUNTER e il rapporto in questione viene trasmesso all'ERM che provvede a svolgere l'elaborazione successiva.

I rapporti che vengono prodotti seguono lo standard COUNTER il quale ha due tipologie di rapporti a seconda delle risorse elettroniche che si stanno prendendo in esame, una per i *journal* e i database (quattro possibili formati di rapporti) e l'altra per gli *e-book* e i *reference work* (sei possibili formati di rapporti). SUSHI 1.0 supporta qualsiasi tipologia di report COUNTER.

Sicurezza

Tutte le transazioni in SUSHI sono fatte usando il protocollo SSL (Secure Sockets Layer)[12]. Questo si traduce in un alto livello di sicurezza delle operazioni, in quanto tutti i dati trasferiti tra *client* e *server* sono criptati e non leggibili dall'esterno. Inoltre, per poter effettuare una richiesta con SUSHI ci si deve autenticare, in quanto il provider richiede *username* e *password* prima di procedere all'elaborazione di una richiesta.

In aggiunta, esiste un ulteriore livello di sicurezza rappresentato dal controllo dell'IP di provenienza della stessa richiesta, al fine di evitare accessi da macchine non autorizzate.

Osservazioni sull'utilizzo di SUSHI

Il beneficio primario dell'utilizzo di SUSHI è l'automazione di processi ripetitivi e molto laboriosi.

Reperire i dati statistici per lo staff di una biblioteca ha sempre richiesto molto tempo, in quanto significava andare sui siti web di ciascun editore e fare il download degli stessi dati. Se si era fortunati, si recuperavano dati che seguivano lo standard COUNTER, ma in molte situazioni ci si poteva trovare di fronte a formati proprietari dell'editore molto difficilmente elaborabili.

Inoltre, in alcuni casi non era prevista neanche la possibilità di effettuare il download dei dati, ma solamente di catturare l'immagine e quindi di poterli avere a disposizione solo in formato cartaceo.

Quindi, SUSHI automatizza e rende possibile tutte queste fasi a cui prima erano preposti i bibliotecari.

Gli editori che ancora non si erano uniformati allo standard COUNTER, per poter essere coinvolti in questo processo si sono dovuti adeguare al nuovo formato standard, ossia COUNTER XML. Perciò, ora le operazioni non solo sono automatiche, ma anche più semplici.

SUSHI, inoltre, permette di poter ottenere dei rapporti a livello consortile, se il consorzio è registrato come un'entità a sé stante.

Non ci sono limiti al numero di richieste che un utente può effettuare e, inoltre, è possibile richiedere intervalli temporali a proprio piacimento, a patto che i dati richiesti siano effettivamente disponibili sui server remoti dei vari editori.

È anche possibile fare in modo che le richieste vengano automatizzate, ossia inviate secondo un calendario prestabilito.

Per fare ciò, è sufficiente fornire al SUSHI Client in maniera preventiva le informazioni relative all'editore di cui si vogliono ricevere i dati statistici, il *customer number* con relativa *password* e una lista di date a partire dalle quali i dati sono disponibili sul server remoto.

Nella descrizione di una transizione SUSHI si è vista la presenza del componente ERM, ma deve essere specificato che non è indispensabile.

I dati che vengono ricevuti possono essere anche caricati in database o software di proprietà della biblioteca che funge da ERM.

Per chi invece possiede l'ERM dell'Innovative Interfaces, ne esiste già una versione che supporta il protocollo SUSHI, così come l'ERM della Ex Libris, Verde.

Bibliografia

- [1] Sito dell'ICOLC:
URL: <http://www.library.yale.edu/consortia/>
- [2] Sito del CILEA:
URL: <http://www.cilea.it>
- [3] Sito del CASPUR:
URL: <http://www.caspur.it/>
- [4] Sito dell'Università degli Studi di Roma 3:
URL: <http://www.uniroma3.it/>
- [5] Sito del consorzio OCUL:
URL: <http://www.ocul.on.ca/>
- [6] Sito del consorzio ANKOS:
URL: <http://www.ankos.gen.tr/index.php>
- [7] Sito del FinElib-Consortium:
URL: <http://www.lib.helsinki.fi/finelib/english/>
- [8] Sito SUSHI:
URL: http://www.niso.org/committees/SUSHI/SUSHI_comm.html
- [9] Sito COUNTER:
URL: <http://www.projectcounter.org/>
- [10] Sito NISO:
URL: <http://www.niso.org/>
- [11] Dal sito di Wikipedia Italia:
URL: http://it.wikipedia.org/wiki/Web_service
Vedi anche l'articolo di A. Manzo "Web Service" su questo numero del Bollettino del CILEA.
- [12] Dal sito di Wikipedia Italia:
URL: http://it.wikipedia.org/wiki/Secure_Sockets_Layer